

# Educational Data Mining for Classification of Students according to their Performance

<sup>[1]</sup> Amruta Joshi <sup>[2]</sup> Prof. A. Thomas

<sup>[1][2]</sup> Computer Science and Engg, G.H. Raisoni College of Engineering, Nagpur.

**Abstract:** - Educational Data Mining is the Data Mining technique that is used to describe a research discipline that uses data collected or gathered from educational areas such as universities and colleges in research that helps to develop a method to gain information or knowledge from the data which we can use to further understand the relationship and environment between student and college or university. Every student lacks at some areas in academic and other performance. As sometimes student don't understand where they are lacking because of which they cannot improve their performance which leads to their poor performance in final results. So there is a need of one system which can help the teachers and students to understand student's performance level.

With the help of this system, Teachers will identify the student's performance level i.e. which student is good in academics which student is weak and can focus more on the weaker students and can take corrective steps from the initial stage to get better result in final exams.

**Keywords:**-- Educational Data Mining.

## I. INTRODUCTION

Educational Data Mining is used to describe a research discipline that uses data from educational settings such as universities and colleges in research that develops method to gain information and knowledge from the data and use it to further understand the student and the university environment.

Educational Data Mining subsequently provides intrinsic knowledge of teaching and learning process to enable better education planning. Educational Data Mining enable faculty members to identify predict and classify students based on their academic performance.

**Educational Data Mining involves some important techniques which are: -**

- A. Prediction
- B. Classification
- C. Data preprocessing
- D. Data rebalancing
- E. Data collection and preparation
- F. Interpretation
- G. Data transform
- H. Data cleaning
- I. Data reduction
- J. Data Integration etc.

*This system mainly focusing on two techniques which are as follows:-*

- A. Prediction and
- B. Classification

### **A. Prediction**

Prediction is a very useful technique in Data Mining which is used to predict a future state using information of past and current state. This technique involves evaluation of the past and current data of the students and based on that data it predicts the final performance of the student so that student can improve their skills before the final exams to get better results. Prediction is the technique which predicts the next or future events by evaluating the available input data.

### **B. Classification**

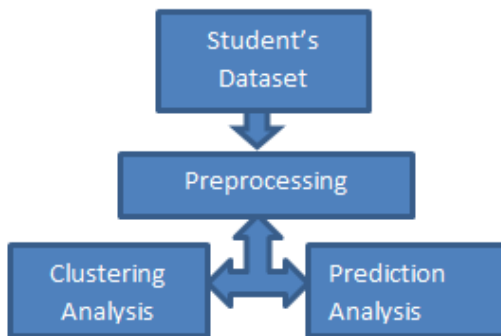
Classification is a two way technique which includes training and testing which are used to map the data into a predefined class. It is a process of supervised learning which separates or divides data into different classes of data sets. It helps to classify the students into different groups or classes according to their performance ratio. By using these two techniques available in the Educational Data Mining, This proposed system will evaluate the past and the current data of the students according to which the future performance of the students can be predicted so that the teachers can identify and evaluate the

performance level of the student and can group them together according to their performance ratio to improve their performance.

In this paper, The former method entails developing a model that can infer a single aspect of the data from some combination of other aspects of the data. The second refers to finding data points that naturally cluster together and can be used to split a dataset into categories. The fuzzy clustering and KFCM were better and the combination between the statistical tools and AI or data mining methods provides better recommendations than using one of them alone. This paper has applied regression analysis as if needs to understand which variable among independent variables are related to the dependent variable and to explore the forms of these relationships.

*There are two analysis techniques which are mainly focused in this paper which are as follows-*

- A. Clustering Analysis
- B. Predicting Analysis



**Fig. 2. System Overview**

**A. Clustering Analysis**

Clustering is one of the main tasks in Educational Data Mining. Clustering divides the data into groups of similar object in which each subgroup is called as cluster which contains all the similar objects of them.

**B. Predicting Analysis**

There are many methodologies which are useful in prediction analysis: such as neural network,

regression, decision trees, logistic regression and generic algorithm etc.

**V. Shaleena K.P, Shaiju Paul “Data Mining Techniques for Predicting Student Performance.”:**

In this paper, they have discussed the classification method for prediction with Decision tree classifiers and methods to solve the class imbalance problem. According to author, Educational data mining is discipline using which we can explore the various kinds of data which is related with student’s education and we can use these methods for understanding students in a better way.

Here, author have discussed that student failure is major and crucial social problem. So, it is very important for educational professionals to understand to reason behind the failure of students. There are two types of classification techniques which are discussed in this paper i.e. White box classification methods and Black box classification methods

White box classification method White box classification is used to explain the predictions by obtaining models using IF – THEN rules. These methods are very easy to interpret. These classification models are used for decision making directly. Black box classification method Black box methods are more accurate but these methods as difficult as compared to White box methods. In this paper, Decision tree classifier have used for students classification according to their data and data re-sampling is used to solve the class imbalance problem. Knowledge discovery and data mining have discussed which generally follows four steps of general framework which are as follows –

1. Data set framework
2. Preprocessing
3. Data mining
4. Interpretation and evaluation

**VI. Krina Parmar Prof. Dineshkumar Vaghela Dr Priyanka Sharma “Performance prediction of students using distributed Data Mining.”:**

In this paper, Classification techniques are used for predicting the student’s performance in distributed environment. According to author, It is very important to generalize the data or information which is contained in those models and to generalize these kind

of rules for global model, specific classifier methods can be used.

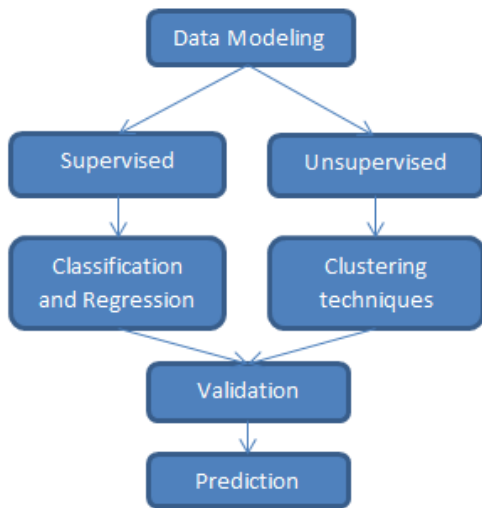
In this paper, two data mining tasks have discussed which are supervised learning and unsupervised learning.

**Supervised Learning**

In this learning method, classification is there and to predict the class, it uses decision tree method and its rule sets. In this, the basic task is classification which is used in the construction of a classifier and this task is very important in data mining. There are various classification algorithms like OneR, NavieBayes, random decision tree, BayesNet, random forest algorithm etc.

**Unsupervised Learning**

In this learning method, clustering and association rule mining are used for finding hidden data or data sets patterns which are ultimately used for decision making from the large data sets. Clustering is used to group the similar kind of data into one cluster and different clusters. There are various clustering algorithms like Simple K-means, Density based clustering algorithm, Hierarchical clustering algorithms etc.

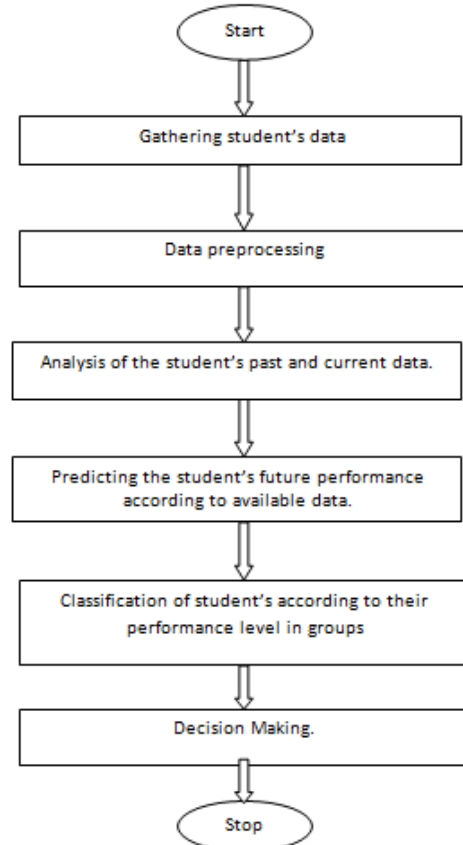


**Fig. 3. Supervised and Unsupervised classification.** Association rule learning method is used to discover the interesting relations between variables in databases.

**III. PROPOSED WORK**

The reference papers which are mentioned above have their work Educational Data Mining and various prediction and classification techniques such as Neural network, Fuzzy logic, Regression technique, Clustering etc. The researcher has used different platforms for the implementation of project. Main software used are the RapidMiner and Matlab tool. And For the implementation purpose need to go with Neural network. The following is the basic block diagram which gives the overall idea of the project.

In this proposed system, the future academic performance i.e. final semester's grades of the students will be predicted according to their past semester's grade points using neural network and then the students will be classified into the four groups like Excellent, Good, Average, Poor.



**Fig. 4. Flow chart of proposed system.** Data Collection

For predicting the student's future performance, their past data needs to be evaluated which includes their all semester's grade points from the beginning of the course i.e. student's grade points from the first semester to the current semester in which they have appeared. Apart from this data the SSC% and HSSC% are also be included.

Using following data of each student and evaluating it, the proposed system can predict the future grade points or performance of the student

This student's data includes:-

- A. SSC grades.
- B. HSSC grades.
- C. All semester grade points of the students.

#### REFERENCES

- [1] Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad : "Educational Data Mining for Prediction and Classification of Engineering Students Achievement," 2015 IEEE 7th International Conference on Engineering Education (ICEED).
- [2] Bo Guo\*, Rui Zhang†, Guang Xu‡, Chuangming Shi§ and Li Yang§ "Predicting Students Performance in Educational Data Mining.": 2015 IEEE International Symposium on Educational Technology .
- [3] Karan Sukhija\* Dr. Manish Jindal Dr. Naveen Aggarwal "The Recent State of Educational Data Mining: A Survey and Future Visions." : 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE).
- [4] Alana M. de Morais and Joseana M. F. R. Araújo Evandro B. Costa "Monitoring Student Performance Using Data Clustering and Predictive Modelling.": IEEE , 2014.
- [5] Shaleena K.P, Shaiju Paul "Data Mining Techniques for Predicting Student Performance." 2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015, Coimbatore, TN, India.
- [6] Krina Parmar Prof. Dineshkumar Vaghela Dr Priyanka Sharma "Performance prediction of students using distributed Data Mining." 2015 IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS)